Ecological Informatics 52 (2019) 46–56

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Classification and regression with random forests as a standard method for presence-only data SDMs: A future conservation example using China tree species

Lei Zhang[a], Falk Huettmann[b], Shirong Liu[c,*], Pengsen Sun[c], Zhen Yu[d], Xudong Zhang[a], Chunrong Mi[e]

[a] Research Institute of Forestry, Chinese Academy of Forestry, Beijing 10091, China
[b] EWHALE LAB, Institute of Arctic Biology, Department of Biology & Wildlife, University of Alaska Fairbanks (UAF), USA
[c] Key Laboratory of Forest Ecology and Environment of State Forestry and Grassland Administration, Research institute of Forest Ecology, Environment and Protection, Chinese Academy of Forestry, Beijing 10091, China
[d] Department of Ecology, Evolution, and Organismal Biology, Iowa State University of Science and Technology, Ames, IA 50011, USA
[e] Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:
Species distribution model
Binary prediction
Numerical prediction
Threshold
Machine learning
Species traits
Climate change

ABSTRACT

The random forests (RF) algorithm is a superb learner and classifier in machine learning applications. This ensemble model is also one of the most popular species distribution model algorithms (SDMs) available to date. RF by default can produce categorical and numerical species distribution maps based on its classification tree (CT) and regression tree (RT) algorithms, respectively. Statistically, CT can also produce numerical predictions (class probability). Many real-world applications (e.g. conservation planning) employ binary presence–absence outputs that use classification thresholds to make these conversions. However, there is little available information regarding the difference in model performance between CT and RT for inference settings. Here, under an ensemble modeling framework, 52 forest tree species with presence-only data for all of China were selected for comparison of the performance of CT and RT algorithms in projecting the distribution and potential range shifts of these species under current and future climates. Five climatic variables were used to develop CT and RT models. Eight threshold-setting approaches were employed to convert numerical predictions into binary predictions. With regard to probabilistic predictions, the relative performance of CT and RT depended on the choice of the evaluation criteria. For both RT and CT, threshold-setting methods significantly altered the determination of thresholds, model performance, and subsequently projections of species range shifts under climate change. The four threshold selection methods (MaxKappa, MaxOA, MaxTSS, and MinROCdist) based on the composite model accuracy measures most often achieved significantly higher model performance than CT default threshold method and other threshold methods. They consistently projected that species' geographical ranges changed in response to climate change with the same direction and magnitude. We argue for choosing RT rather than CT as the SDM if model discrimination capacity (the ability to differentiate between occurrences of presence and absence) is viewed as more important than model reliability (the agreement between predicted relative indexes of occurrence and observed proportions of occurrence), and vice versa. In line with gradient theory, we can recommend the use of numerical predictions for species distribution modeling since they help to convey more information than binary predictions. Binary conversion of model outputs should only be carried out when it is clearly justified by the application's objective. The four aforementioned threshold methods are promising objective methods for binary conversions of continuous predictions when presence-only data are available. This study proposes guidelines on how machine learning can be used for specific applied and theoretical applications in a SDM context.

* Corresponding author: Research Institute of Forestry, Chinese Academy of Forestry, No.1 Dongxiaofu, XiangShan Road, Haidian, Beijing 100091, China.
E-mail addresses: lei.zhang@caf.ac.cn (L. Zhang), fhuettmann@alaska.edu (F. Huettmann), liusr@caf.ac.cn (S. Liu), zyu@iastate.edu (Z. Yu).

https://doi.org/10.1016/j.ecoinf.2019.05.003
Received 14 January 2019; Received in revised form 30 April 2019; Accepted 2 May 2019
Available online 07 May 2019

1. Introduction

Species distribution models (SDMs) have been increasingly applied to tackle a wide range of questions in ecology, evolution, biogeography, forestry, climate change, and conservation biology. For instance, they have been used to quantify the environmental niche or niche shifts of species (e.g. Baltensperger et al., 2017; Han et al., 2018; Petitpierre et al., 2012), to identify hierarchies of environmental drivers (e.g. Crimmins et al., 2011; Zhang et al., 2016), and to generate and test biogeographical and ecological hypotheses (e.g. Patsiou et al., 2014; Zhang et al., 2016). SDMs have also been used to inform the prospective design of surveys for rare species (e.g. Sarre et al., 2013) and to map suitable sites for species recovery and reintroduction (e.g. Bleyhl et al., 2015; Kandel et al., 2015). Others have used SDMs to assess the effects of environmental changes on species distribution (e.g. Nenzén and Araújo, 2011; see https://www.fs.fed.us/nrs/atlas/ for large-scale application) and for modeling of species assemblages (biodiversity) by stacking individual species predictions (e.g. Cao et al., 2013; Cooper and Soberón, 2018; Gavish et al., 2017), or for landscape management (Drew et al., 2011). Booth et al. (2014) and Guillera-Arroita et al. (2015) reviewed and detailed the current use of SDMs, whereas the pure data mining concept is widely ignored still (but see work by Craig and Huettmann, 2009; Huettmann and Ickert-Bond, 2017).

Machine learning is a growing and leading approach for a wide range of modern analysis problems including classification, regression, data mining and predictions (Mueller and Massaron, 2016). It has developed into its own type of research and carries much literature and code. Due to its magnitude and depth, for now, machine learning applications still lag behind though when it comes to natural resource management applications (but see Cushman and Huettmann, 2010; Drew et al., 2011).

Random forests (RF), one of the most popular SDM algorithms in those applications (Cutler et al., 2007; Oppel et al., 2012; Prasad et al., 2006), is a deep analysis platform (Breiman, 2001a, 2001b) and has the flexibility needed to perform several types of statistical data analysis, including regression, classification, survival analysis, and unsupervised learning (Breiman, 2001a; Breiman and Cutler, 2004; Liaw and Wiener, 2002). It's a major 'learner' and classifier in machine learning (Fernández-Delgado et al., 2014; Hastie et al., 2009). Besides other concepts to obtain best bagging tree output, the model is based on unpruned CARTs (Breiman et al., 1984) and those can be categorized as a classification (and regression) tree (Strobl et al., 2009). This type of classification tree explains the variation of a single response variable by recursively splitting the data in a binary fashion into progressively more homogeneous groups, using one or more explanatory variables (Breiman, 2001a; Breiman and Cutler, 2004; Liaw and Wiener, 2002). The response variable is usually either numeric (regression trees) or categorical (classification trees), and the explanatory variables can be numeric and/or categorical (De'ath and Fabricius, 2000). Owing to its hierarchical nature, a RF is capable of capturing non-linear and correlated relationships in predictor variables. It can be particularly useful for inference from complex data based on predictions (Breiman, 2001b). A major application is species distribution modeling, which often exhibits complex interactions in predictor variables. Due to bagging, the Law of Large Numbers (Feller, 1968) shows that RF does not tend to overfit data (Breiman, 2001a, 2001b). A RF model is expected to balance the accuracy and robustness of predictions because it inherently incorporates the concept of bagging and ensemble learning (Hastie et al., 2009). Therefore, RF models often perform better than other SDMs (e.g. Cutler et al., 2007; Jafariana et al., 2019; Mi et al., 2017; Peters et al., 2007; Prasad et al., 2006; Zhang et al., 2016). Inference is drawn from the prediction (Breiman, 2001a, 2001b). Based on its underlying classification (CT) and regression (RT) tree algorithms, RF can produce categorical/classified/binary and numerical predictions, respectively. RF model applications in the form of CT and RT models are commonly and successfully used in species distribution modeling with species' presence/absence data (Cutler et al., 2007; Mi et al., 2017; Peters et al., 2007; Zhang et al., 2014). Statistically, CT can also produce numerical predictions (class probability or better called species relative index of occurrence, Liaw and Wiener, 2002). However, researchers have often focused on the application of either CTs (e.g. Mi et al., 2017; Peters et al., 2007) or RTs (e.g. Gavish et al., 2017; Kandel et al., 2015; Nenzén and Araújo, 2011; Zhang et al., 2016) in a SDM context. Required knowledge on the difference in model performance between CTs and RTs remains unclear. It is also worthwhile to explain that the RF code comes from Breiman (2001a) and then was released and implemented in several versions. However, the regression solution was not really part of Leo Breiman's initial code and mostly is based on Andy Liaw's implementation in R (Liaw and Wiener, 2002).

SDMs are usually constructed through a series of methods that relate a set of environmental predictors with information on species distributions (Drew et al., 2011; Guisan and Zimmermann, 2000). Information about the distributions of species, frequently housed in museum and herbarium collections, atlases, plant lists, or provided by volunteer observation networks (citizen science), is becoming increasingly available over the internet (Graham et al., 2004; Huettmann and Ickert-Bond, 2018). These data sets are typically composed of 'presence-only' (i.e. no information is usually available on the absence of most species), presence–absence or abundance data (Mateo et al., 2010). Accordingly, SDMs can be categorized in two groups: models that only need presence data (profile techniques) vs. those that require both presence and absence data or that require abundance data (group discrimination techniques; Mateo et al., 2010). The application of data mining is useful to either of those approaches as it is known to resolve distinctions from presence vs. absence or random well. Data mining, using machine learning and specifically RF, is the method of choice to find 'a signal', any signal as well as outliers, in data (e.g. Fernández-Delgado et al., 2014; Mi et al., 2017). Presence-only SDMs are more likely to yield potential distributions or the fundamental niche information for a species, whereas presence–absence SDMs are more likely to reflect the natural distribution or realized niche of a species (Zaniewski et al., 2002). Either way, such concepts most often yield the best-possible solution when high-powered algorithms are employed (Elith et al., 2006; Fernández-Delgado et al., 2014). Comparisons of various SDMs indicate presence–absence models tend to perform better than presence-only models (Elith et al., 2006; Mateo et al., 2010; Oppel et al., 2012). Thus, presence–absence models are increasingly used when only presence data are available, by creating artificial absence data (i.e. pseudo-absence data; Zaniewski et al., 2002; Mateo et al., 2010; Barbet-Massin et al., 2012). Several studies have suggested that pseudo-absence data should be restricted to locations that are documented to be distinctly unsuitable for the occurrence of a particular species (Mateo et al., 2010; Zaniewski et al., 2002). But once more, in real-world data mining, those are smaller differences as the latter methods can find the signal in a rather reliable fashion (e.g. Craig and Huettmann, 2009).

When developing SDM models, presence–absence data (response variables) are often treated using numeric (usually $Y = 1$ for presence and $Y = 0$ for absence) or categorical (binary value) variables. Correspondingly, they result in model prediction outputs that provide a value of the relative occurrence index scaled from 0 to 1, or a binary value represented by presence and absence (see the detailed description in the BIOMOD manual, Thuiller et al., 2009). However, in resource management, climate change and environmental conservation applications (e.g. reserve design, biodiversity assessment, climate change), information that is presented in a binary format such as species presence/absence may have more practical applications than that presented as a continuous index (Baltensperger et al., 2017; Fernandes et al., 2018; Kandel et al., 2015). Therefore, a threshold is needed to convert continuous indexes to binary presence–absence predictions. Furthermore, many commonly used performance measures such as the true skills statistics (TSS) and Kappa require binary data (Fielding and

Bell, 1997; Pearce and Ferrier, 2000). Although many threshold selection methods exist for presence/absence data (Jiménez-Valverde and Lobo, 2007; Liu et al., 2005; Nenzén and Araújo, 2011), very few methods have been yet proposed for use with presence-only data (Liu et al., 2013; Liu et al., 2016), and it is somewhat unclear which threshold method is most appropriate for CTs and RTs.

In view of the above here we raise the following questions about CT and RT predictions for a valid inference: Which concept works best to predict species distribution with presence and pseudo-absence data; and which threshold is most appropriate if binary conversion is wanted or necessary?

While there are many implementations of the RF base code (e.g. in R, Python, Fortran, SPM software of Salford Systems Ltd) with different strengths and weakness (Liaw and Wiener, 2002; Strobl et al., 2009; Ishwaran and Kogalur, 2007; Brieuc et al., 2018; see Herrick, 2013 for an assessment), here we choose the R package "randomForest" (Liaw and Wiener, 2002) to construct CT and RT prediction models. We did so because it is particular popular in SDM work with an open- access and -source characteristic, and because R has a large library of statistical packages relevant to SDMs for data preprocessing and post processing (e.g. Freeman and Moisen, 2008a; Hijmans, 2012; Thuiller et al., 2009); those can easily be linked.

Under an ensemble modeling framework, 52 forest tree species from China were selected for comparison of the performance of CT and RT algorithms in projecting the distribution and potential range shifts of these species under current and future climates. This matters a lot because in China, concern on environmental protection and forest resource conservation has prompted the rapid development of tree plantations (Bryan et al., 2018; Li, 2004). Currently the area of forest plantations stands in China at 69 million hectares, or 24.8% of the total area worldwide (277.9 million hectares), which is well ahead of any other country (FAO, 2015). Moreover, in joining the international efforts of mitigating global climate change, China has set a target to increase forest carbon sink by expanding forest cover as a key measure in the forestry *sec*tor (26% forest cover by 2050, and 40 million new hectares by 2020 when compared to 2005 levels; SFAC, 2010). For achieving the goal of expanding forest cover, large areas of new plantations need to be established. China has a land area of 9.6 million square kilometers, spans a large range of climate and nature environments (Song and Zhang, 2010). The identification of climate requirements and predictions of potential range shifts of native tree species under an altered climate will greatly facilitate the assumed success in establishing new plantation forests. We also anticipate that this study will provide a scientific basis for the choice of species and sites for the large-scale forestation practice. Finally, these steps are facilitated by randomForest and here we assess whether that concept and its workflow can stand as a generic template.

2. Materials and methods

Fig. 1 shows the overall work flow for our ensemble forecasting approach. It's meant to be a generic concept to be applied to virtually any model prediction question with presence only data.

2.1. Data sources

Fifty-two native forest tree species that occur in China were selected for a comparison of the performance of CT and RT algorithms. The distribution datasets for these 52 tree species were originally derived from the 1:1,000,000 Vegetation Distribution Map of China (EBVMC, 2001). Those were obtained from the Environmental and Ecological Science Data Center for West China of the National Natural Science Foundation of China (http://westdc.westgis.ac.cn; see http://www.nsii.org.cn/mapvege for raw data maps). These distribution datasets were then resampled to a spatial resolution of 8 km. See Table S1 in Research Data (Mendeley Data) for the ecological requirements, biological characteristics and conservation status of these 52 tree species.

We started to characterize the environments in China based on 19 biologically relevant proxy climatic variables (Table 1) drawn from the WorldClim data set at a resolution of 30 arc seconds (www.worldclim.org). Baseline climatic data were obtained from the average of the period 1960–1990, and these data were rasterized to a cell size of 8 km. RandomForest does not suffer much from correlations (Breiman, 2001b; Cutler et al., 2007; De'ath and Fabricius, 2000; Herrick, 2013). But to address the argument completely and to reduce the risk of multicollinearity, only variables with a Pearson's correlation coefficient < 0.80 were used for this study. We finally kept the following five climatic variables for CT and RT models: annual mean temperature, annual range of temperature, isothermality, annual precipitation, and precipitation seasonality (coefficient of variation).

To assess RF model transferability related to forthcoming future projections of climate change, four greenhouse gas scenarios (i.e. four representative concentration pathways (RCPs) were used: RCP2.6, RCP4.5, RCP6.0, RCP8.5) as well as three global climate models (BCC-CSM1-1, CCSM4 and MRI-CGCM3). Climate change scenarios were averaged for the 20-year period: 2061–2080 (2070s). For the future climatic projections, the same set of five climate variables were obtained from the WorldClim data set for all 8 km × 8 km grids.

2.2. Generation of pseudo-absence

To control for a standardized research design and a more generic assessment of our work, beyond China, the following two approaches for RF were used to randomly select pseudo-absences, as recommended by Barbet-Massin et al. (2012):

(1) Environmentally stratified sampling. The locations, where all predictor variables fall within the extreme values (both maximum and minimum limits of each predictor) as determined by species presence sites, were defined as areas suitable for the occurrence of a particular species. The remaining locations were termed as 'potential' absences. This process was implemented by the surface range envelop model (SRE) in the BIOMOD2 package (Thuiller et al., 2009) in the R platform (hereafter, the 'SRE' method).
(2) Geographically stratified sampling. Any point located at least two degrees in latitude or longitude from any presence point were selected as 'potential' absences (the '2 degree' method). It assumes that the closer a location is to a known presence point, the more likely it will be that the species will be found.

2.3. The randomForest model (in R)

In RF models, bootstrap samples are drawn from rows and predictors (= bagging) to construct multiple decision trees, and each tree is grown with a randomized subset of predictors, hence the name "random" forests (Breiman, 2001a). Hundreds to thousands of trees are grown (a "forest" of decision trees). In a typical bootstrap sample, approximately 2/3 of the original data are used to build the model with the remaining 1/3 are held in reserve (out-of-bag OOB data) (Breiman and Cutler, 2004). Once the tree is grown with the bootstrap sample, it can further be used to predict the OOB data, the misclassification (error) rate averaged over all trees is called OOB error which is an unbiased estimate of RF generalization error. In RF, the trees are fully grown without pruning, and they are used to predict new data by aggregating the predictions of the trees (i.e. proportion votes for classification, average for regression; Liaw and Wiener, 2002). In a typical CT, the resulting model output is categorical, and the 'winning' class for an observation is the one with the maximum ratio of proportion of votes (Default is 1/k where k is the number of classes). For presence–absence data, the ratio of proportion of votes for presence or absence ranges from 0 to 1, and sum of ratios for both them is equal to 1. As such, the resulting ratio for presence in CT could been taken as a relative index of

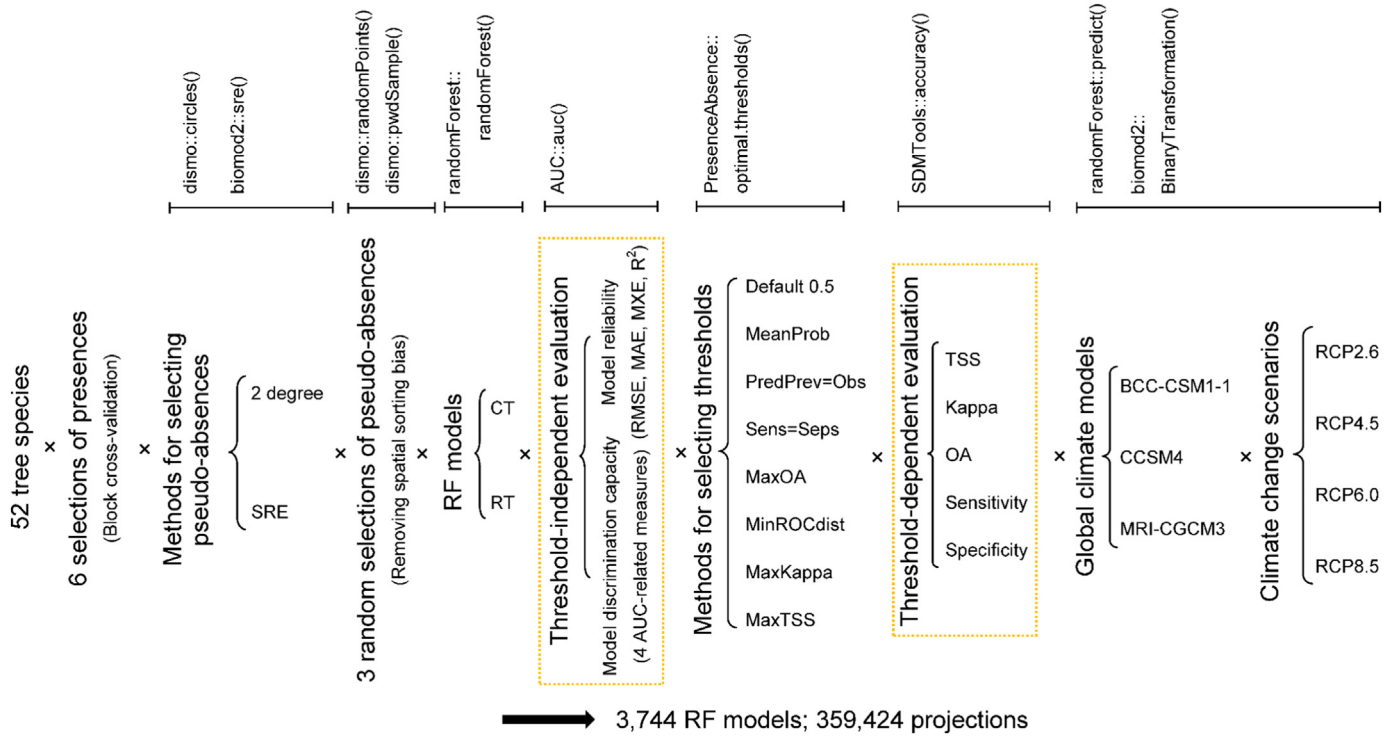

Fig. 1. General framework for ensemble modeling and R functions used in this study.

Table 1
Biologically climatic variables.

Code	Variable
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp – min temp))
BIO3	Isothermality (BIO2/BIO7)
BIO4	Temperature Seasonality
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

occurrence (numerical prediction) (Strobl et al., 2009). For CT, the default threshold cutoff used to determine either presence or absence is 0.5, and CT considers the absences and presences as mutually exclusive. When RT is used for presence–absence data, RF produces a habitat-suitability (numerical) map, where suitability values range from 0 to 1. The numerical predictions of both CT and RT are converted to binary predictions of species presence and absence through threshold methods (see below).

In this RF implementation two parameters can be manipulated, i.e. the number of trees grown, and the number of variables to try at each split. By default, both CT and RT are run with 500 decision trees to find an optimum; and at each split, the square root and 1/3 of the total number of predictors are used to identify the best split for CT and RT, respectively (Liaw and Wiener, 2002). Node-splitting is based on discerning the predictor that minimizes the within group variance for CT, or error when regressed against the response for RT. Here we adopted the default R memory allocation for data analysis. Our purpose was to investigate the relative performance of CT and RT, and the difference among threshold approaches. Because CT and RT both often perform better than other SDMs with default parameters settings (e.g. Cutler et al., 2007; Herrick, 2013) and the outcome of both of them is not very sensitive to modifications of these parameters (Brieuc et al., 2018; Cutler et al., 2007; Peters et al., 2007), it was not deemed necessary to fine-tune all the RF models to their optimal capacity. Modeling was carried out in the R 3.3.3 statistical programming environment (R Core Team, 2017) using the package known as 'randomForest' (Liaw and Wiener, 2002) which is based on Breiman and Cutler (2004) and which includes a regression extension not done or really approved by Breiman (2001a). The fine-tuned commercial code from Breiman is with Salford Systems Ltd. (https://www.salford-systems.com/).

2.4. Ensemble modeling of species distributions

Random data splitting (cross-validation) does not provide truly independent validation when a dependence structure is present (Hijmans, 2012). Here we applied a block cross-validation to the presence data following the approach of Roberts et al. (2017), where the presence data area was divided into four geographically non-overlapping areas to split the data into blocks rather than random assignment of locations to a split. The block cross-validation could increase spatial independence of training and testing datasets, and help to evaluate model transferability rather than just it interpolation accuracy (Roberts et al., 2017). Presence records are split into two sets based on their longitude using a meridian as a dividing line. Then, these two halves (with the same longitudinal range) are separately split into two equal parts (with the same latitudinal interval) using parallels. Each two blocks were used in turn as model-training data, while the two others were separately used to test model and select the optimal threshold. Using the 2 degree and SRE methods, RF here samples pseudo-absence data from the entire study region.

By combining the block cross-validation strategy with the pairwise

distance sampling method proposed by Hijmans (2012) to select the pseudo-absence points for the model test and threshold selection sets, spatial sorting bias was removed and thus the effect of spatial autocorrelation on the performance evaluation suppressed. In the model-building process, we also kept the ratio between the number of presences and absences in the calibration and testing dataset constant at 1:1. This is a recommended method used to find the optimal transforming threshold (Liu et al., 2005) and to achieve the highest model accuracy (Barbet-Massin et al., 2012; Evans et al., 2011; Freeman and Moisen, 2008b) when using RF and a presence/pseudo-absence dataset to develop SDMs. Because chance plays a part in the choice of the pseudo-absences, and we provide here a global model method, we repeat this procedure three times, independently. This was done in an effort to reduce variability in the model-building process and subsequent predictions. Thus, 18 different CT and RT models were calibrated for each species with the pseudo-absence selection methods (Fig. 1). With 18 SDMs, four gas emission scenarios and three global climate models, we obtained an ensemble of 216 projections for each of our species under future climate.

2.5. Threshold selection in aspatial metrics

Eight prevalent threshold approaches were used for optimal threshold determinations and subsequent binary conversions. These analyses were conducted using the R package 'PresenceAbsence' (Freeman and Moisen, 2008a):

(1) Default 0.5: Taking a fixed value, default 0.5, as the threshold.
(2) MeanProb: Taking the average predicted probability of the threshold-selecting data as the threshold.
(3) PredPrev = Obs: The threshold where the predicted prevalence (the proportion of sites occupied) is equal to the observed prevalence of threshold-selecting data.
(4) Sens = Seps: The threshold where sensitivity (the proportion of observed presences correctly predicted as presence) equals specificity (the proportion of observed pseudo-absences correctly predicted as pseudo-absence) for the threshold-selecting data.
(5) MaxOA: The threshold that results in the maximum value of overall accuracy (OA) for the threshold-selecting data. OA measures the proportion of correctly classified presences and absences.
(6) MinROCdist: The threshold corresponds to the point on receiver operating characteristic (ROC) curve (sensitivity against 1-specificity) which minimize the distance to the top-left corner (0,1) in the ROC plot. The area under curve (AUC) of the ROC is a threshold-independent model evaluation indicator and is also independent of both species prevalence and classification threshold (Fielding and Bell, 1997).
(7) MaxKappa: The threshold that results in the maximum value of kappa for the threshold-selecting data. Kappa measures the extent to which the agreement between observed and predicted is higher than that expected by chance alone.
(8) MaxTSS: The threshold that results in the maximum value of the true skill statistic (TSS) for the threshold-selecting data. TSS = sensitivity + specificity − 1. TSS has all of the advantages of Kappa but is not sensitive to prevalence (Allouche et al., 2006).

2.6. Model evaluation

2.6.1. Threshold-independent evaluation

For numerical prediction relating to CT and RT, the predictive performances were evaluated using root mean square error (RMSE), mean absolute prediction error (MAE), coefficient of determination (R^2), mean cross entropy (MXE), and area under the curves (AUCs) of four threshold-independent measures: the area under the sensitivity curve, the area under the specificity curve, the area under the accuracy curve, the area under the ROC curve. The latter four measures related to AUC were estimated using the 'AUC' package in R statistical environment (Ballings and Van-den-Poel, 2013). Measures of AUC avoid the need to choose a threshold value that separates presence from absence (i.e., it is threshold independent), and furthermore it describes the overall ability of the model to discriminate between two cases.

The RMSE, MAE, R^2, and the MXE were calculated for the dataset as Liu et al. (2011):

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(p_i - o_i)^2}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}|p_{\mathrm{i}} - o_i|$$

$$R^2 = 1 - \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2/[p(1-p)]$$

$$MXE = -\frac{1}{n}\left[\sum_{o_i=1}\ln p_i + \sum_{o_i=0}\ln(1-p_i)\right]$$

where p_i and o_i are the predicted and observed values (1 for presence and 0 for pseudo-absence) for site i, $\bar{o}$ is the mean of observed values, n is total number of sites, p is the observed prevalence of model-testing data.

2.6.2. Threshold-dependent evaluation

We quantified the accuracy of binary maps produced by threshold approaches using five measures of accuracy: Kappa, TSS, OA, sensitivity and specificity. Kappa, TSS and OA are composite measures of model performance as they attribute different weights to the various types of prediction errors (e.g. omission, commission, or confusion).

2.7. Additional statistical assessment methods employed

To distinctively characterize the spatial differences among binary predictions for each split-sample, the current analysis focused on the sites where each species was determined to be present by any one of the eight threshold approaches. The spatial correspondence among predictions produced by the eight threshold-setting approaches was quantified by using principal component analysis (PCA; Nenzén and Araújo, 2011). The first axis of the PCA, which accounts for a proportion of the variance among predictions and captures consistent patterns in species distributions, was kept as a synthetic variable describing the spatial correspondence for each split-sample. If there was no variability across predictions, the first consensus axis would explain 100% of the variation. To accomplish this, we converted a binary map into a continuous value with 1 representing presence and 0 representing absence. Regarding the binary prediction maps, we also identified the pairwise difference among threshold approaches for each split-sample with the McNemar test (a statistical test used on paired nominal data) in R at the 0.05 significance level. The frequency of the McNemar tests that indicated no significant pairwise difference existed among threshold approaches in ensemble predictions (projection) and was used to characterize the spatial correspondence in binary maps for each species.

The Friedman rank sum test (a non-parametric test on the differences of several related samples), along with a post hoc Nemenyi test (pairwise test for multiple comparisons of mean rank sums), were used to compare threshold cutoffs (model accuracy) and species range shift projections between threshold methods, while a Wilcoxon signed-rank test was used to compare model accuracies between CT and RT.

3. Results

3.1. Differences in model accuracy for numerical predictions

We found that the relative performance of CT and RT depended on

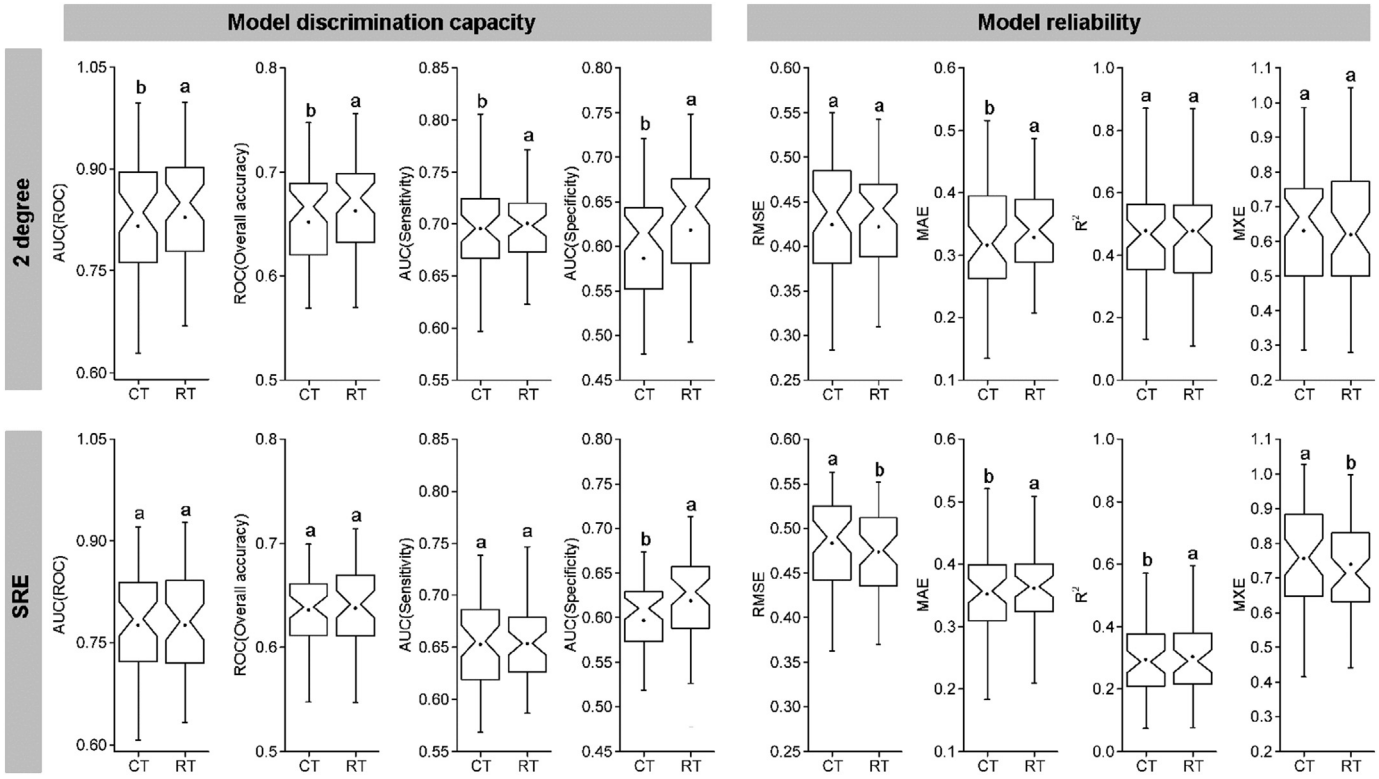

Fig. 2. Box-whisker plots of differences in model accuracy between random forests regression (RT) and classification (CT) algorithms when data were pooled for all species. Boxes were notched at the median, with 95% confidence intervals. Dots show the mean value across all species. Different letters indicate significant differences according to Wilcoxon signed-rank test ($P < .05$).

the choice of evaluation criteria (Fig. 2). For 2 degree method, Wilcoxon signed-rank tests indicated AUC values of ROC, OA, sensitivity, and specificity were significantly higher for RT than that for CT. When MAE was used to assess model performance, CT performed better than RT. CT and RT performed equally well with respect to RMSE, R^2, and MXE. For SRE method, RT performed better than CT in terms of RMSE, R^2, MXE, and AUC value of specificity. MAE was significantly higher for RT than that for CT, while there was no significant difference between CT and RT for other accuracy measures.

3.2. Variation in aspatial model accuracy for binary predictions

For both CT and RT, binary prediction accuracies of the four threshold approaches (MaxKappa, MaxOA, MaxTSS, and MinROCdist) were often higher than that of the other threshold methods (Default, MeanProb, Sens = Spec, and PredPrev = Obs; Fig. B1 in Mendeley Data); we therefore only present the results for the four threshold methods in CT and RT along with the CT default method.

In most combinations of RF models (CT or RT) and the pseudo-absence method (SRE or 2degree) the four thresholds selection methods outperformed the default 0.5 threshold of CT (with few exceptions for specificity). However, we did not observe any significant differences between any of the four thresholding methods, within and between RF methods (Fig. 3).

3.3. Inconsistency in optimal thresholds

With respect to the SRE and 2 degree methods, the Friedman rank sum tests indicated there was a significant difference in the optimal threshold among the eight threshold methods for both CT and RT, except for CT when the SRE method was used for pseudo-absence selection (Fig. B1 in Mendeley Data).

In terms of the optimal threshold, the CT default method did differ significantly from the four aforementioned threshold methods in CT and RT, while the four threshold methods did not differ significantly within and between CT and RT for the 2 degree method (Fig. 3). For the SRE method, the CT default method did not differ significantly from the four aforementioned threshold methods in CT, however, it was all significantly higher than the four threshold methods in RT except for MinROCdist. The four threshold methods did not differ significantly within CT and RT for the SRE method.

3.4. Discrepancy in range change projections

For 2 degree and SRE methods, the Nemenyi multiple comparison tests showed that the four threshold methods did not differ significantly within and between CT and RT in terms of most of the species range shift indicators (with few exceptions for habitat gained and lost with 2 degree pseudo absences and for habitat gained with SRE pseudo-absences) (Table 2).

But when it comes to total habitat area, the CT default threshold method differed significantly from the four threshold methods in CT and RT for the 2 degree method (Table 2). However, the CT default method did not differ significantly from the four threshold methods in CT and RT for most of the other species range shift indicators (with few exceptions for habitat area related indicators). This was also the case for the SRE method for all species range shift indicators except for total habitat area and habitat gained.

3.5. Spatial uncertainty in binary predictions

Substantial variability was observed in binary predictions among threshold methods (Fig. 4). When averaging the map correspondence (estimated by PCA) from split-sample bouts for each species, Wilcoxon signed-rank tests indicated that CT did not differ significantly from RT in terms of spatial correspondence when the SRE method was used to

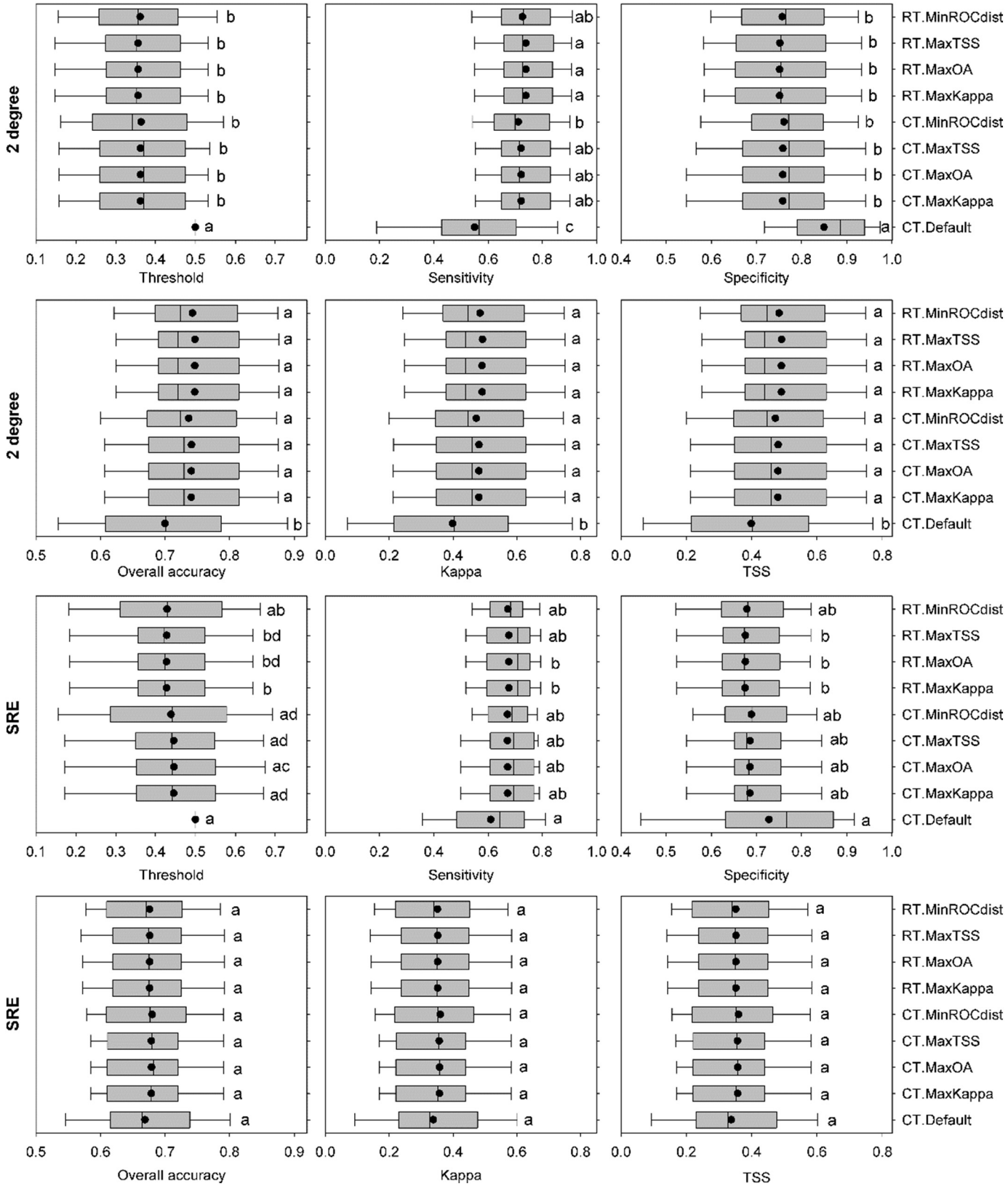

Fig. 3. Differences in threshold cutoff and model accuracy among threshold approaches. Dots show the mean value across species. Section 2.5 provides the abbreviations for acronyms used to name the threshold approaches. Note: SRE, surface range envelop model.

select pseudo-absences. Meanwhile, CT was significantly higher than RT under current and future climates when the 2 degree method was used to select pseudo-absences (Fig. 4).

The McNemar tests indicated the four aforementioned threshold methods most often yielded a binary map without pairwise significant difference; and the pairwise significant difference most frequently

Table 2
Difference in the changes in the range of species (relative to baseline) predicted by classification (CT) and regression (RT) algorithms of random forests.

	Range shift	CT					RT			
		Default 0.5	MaxKappa	MaxOA	MaxTSS	MiniROCdist	MaxKappa	MaxOA	MaxTSS	MiniROCdist
2 degree	Total habitat area (km^2)	a	b	b	b	b	b	b	b	b
	Total range change (%)	a	ac	ac	ac	ac	bc	ac	bc	ac
	Habitat gained (%)	a	bc	bc	bc	b	ac	ac	ab	ab
	Habitat lost (%)	a	ab	ab	bc	ac	b	b	b	b
	Eastward shift (m)	a	a	a	a	a	a	a	a	a
	Northward shift (m)	a	a	a	a	a	a	a	a	a
	Uphill shift (m)	a	a	a	a	a	a	a	a	a
SRE	Total habitat area (km^2)	a	bc	bc	bc	ab	b	b	b	ac
	Total range change (%)	a	a	a	a	a	a	a	a	a
	Habitat gained (%)	a	bc	bc	b	b	ac	a	a	a
	Habitat lost (%)	a	a	a	a	a	a	a	a	a
	Eastward shift (m)	a	a	a	a	a	a	a	a	a
	Northward shift (m)	a	a	a	a	a	a	a	a	a
	Uphill shift (m)	a	a	a	a	a	a	a	a	a

Methods in the same row with the same letters are not significantly different at $P \leq .05$ according to the Friedman rank sum test, along with a post hoc Nemenyi test. Section 2.5 provides the abbreviations for acronyms used to name the threshold approaches. Note: SRE, surface range envelop model.

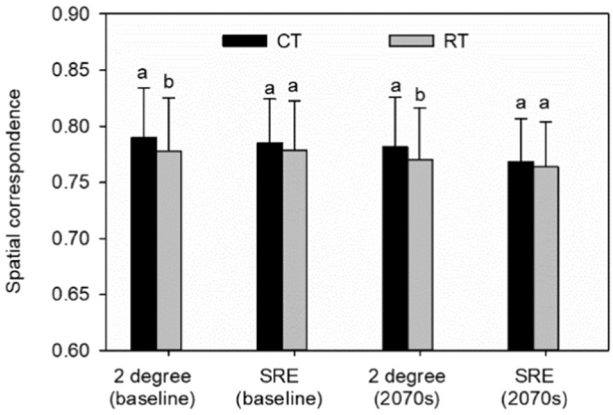

Fig. 4. Spatial correspondence (as judged by the first axis of principal component analysis) among binary predictions produced by eight threshold approaches when data were pooled for all species. Different letters indicate significant differences ($P < .05$) Note: SRE, surface range envelop mode.

occurred between other approaches and the four threshold approaches or the default 0.5 approaches; intermediate frequency of pairwise significant difference often occurred among the other approaches under current and future climates (Fig. B2 in Mendeley Data).

When the four aforementioned threshold methods and the CT default method were compared pairwise within and between CT and RT for both the 2 degree and SRE methods, the four threshold methods within CT and RT more frequently yielded consistent binary maps (without pairwise significant differences; Fig. 5). Meanwhile, a pairwise significant difference was almost always observed between CT and RT, and between the CT default method and four threshold methods from CT and RT.

4. Discussion

This is the first study to compare the model performance of RT and CT using different thresholding methods. Our model predictions have a good to very good accuracy and allow to tackle questions of modern ecological applications, e.g. selection of sites and species for reforestation with future climate change in mind, and overall tree species conservation. Further, we applied one of the best algorithms (RF) known for the SDMs on a national level in an open source code with open access data. As shown by Regmi et al. (2018) we think that it presents a paradigm shift in the sciences, and those methods become a standard for such questions.

In statistical terms we found that the relative performance of CT and RT depended on the choice of the evaluation criteria (Fig. 2). This phenomenon might be ascribed to the fact that different composite measures are based on different algorithms and assumptions (Liu et al., 2011), e.g. attributing different weights to the various types of prediction errors (e.g. omission, commission, or confusion). In short, different measures of accuracy have different strengths and weaknesses (Fig. B1 in Mendeley Data), and no measure provides a universal rating for SDM performance. Moreover, for numerical predictions, measures of accuracy often characterize two aspects of SDMs: discrimination capacity and reliability (Liu et al., 2011). Discrimination capacity (e.g. AUC-related values) measures the ability to differentiate between occurrences of presence and absence based on model predictions. Reliability (e.g. RMSE, MAE, R^2, MXE) tells us about how closely predicted relative occurrence indexes match observed proportions of occurrence, or goodness of fit. The relative importance of reliability and discrimination capacity depends on the use of the SDM and the experience of the user (Pearce and Ferrier, 2000). In practice, the selection between CT and RT depends on the specific species, model accuracy measures, and expert experience (Liu et al., 2011). For instance, if SDMs are used to estimate the total population size for a species by predicting the probability of the species occurring at a large set of sites within a region, model reliability should be viewed as more important than model discrimination capacity (Pearce and Ferrier, 2000). When SDMs are used to identify potential re-introduction sites for endangered species, more attention should be paid to model discrimination capacity (Pearce and Ferrier, 2000).

When converting numerical predictions into binary predictions, the optimal threshold varies with the choice of threshold-setting methods. Our study showed that the choice of thresholds had practical consequences for estimating of model performance and species range shifts under climate change (Fig. 3, Table 2). This was supported by previous research studies (Cao et al., 2013; Freeman and Moisen, 2008b). Hence, the use of an appropriate threshold appears to be a better choice for binary conversions for RF. The threshold approach of MaxTSS was of particular interest because it was the least sensitive to modeling method. It was proven mathematically and demonstrated empirically that the threshold method based on maximizing the sum of sensitivity and specificity (equivalent to MaxTSS) produces the same threshold using either presence/absence data or presence-only data (random points are used as true absences) (Liu et al., 2013). This was supported by Liu et al. (2005) and Jiménez-Valverde and Lobo (2007), who found the sensitivity-specificity sum maximization approach produced the most accurate predictions. Recently, Liu et al. (2016) further indicated MaxTSS produced similar results (optimal threshold and model

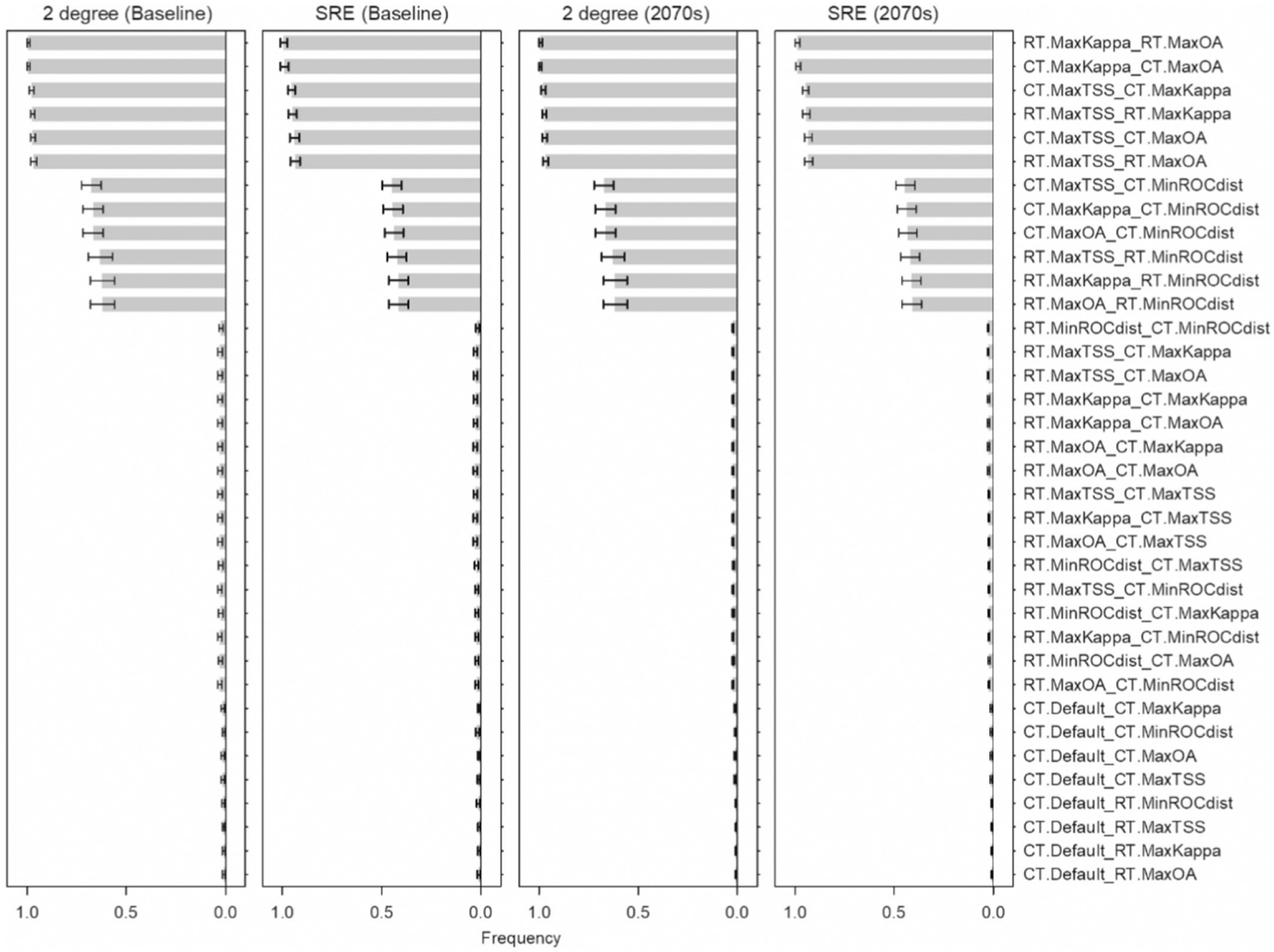

Fig. 5. The frequency of spatial correspondence (as judged by McNemar tests) for pairwise among threshold approaches when data were pooled for all species. Data are presented as mean (gray bar) with 95% confidence interval (black bar). Section 2.5 provides the abbreviations for acronyms used to name the threshold approaches. Note: SRE, surface range envelop model.

accuracy) for RF when using either presence/absence or presence-only datasets. This fact was also noted during our analysis; MaxTSS was among the top four approaches (MaxKappa, MaxOA, MinROCdist, and MaxTSS). We further indicated that the top four approaches performed equally in terms of model performance, threshold determination, and range shift projection, and they all often performed better than the four other approaches. Based on this finding, we can infer that the top four methods can produce the same threshold using either presence-only data or presence/absence data for CT and RT models. Therefore, they can be considered as promising threshold methods for RF when only presence data are available.

The discrepancy in relative performance among threshold approaches between our study and other researchers (Jiménez-Valverde and Lobo, 2007; Liu et al., 2005; Norris, 2014) is probably caused by the difference in model-testing datasets, SDM classes, and species traits. In the present study, a block cross-validation was used to evaluate model performance and determine threshold cutoffs, while the other studies usually derived a threshold based on model-building datasets (Jiménez-Valverde and Lobo, 2007; Liu et al., 2005). In addition, the choice of SDM model may also influence the values of the relative index of occurrence, and thus the optimal threshold for different models would vary (Nenzén and Araújo, 2011). Species traits can substantially influence the determination of optimal thresholds. Species with prevalence values closer to 0.5 generally produced better predictions than scores more distant from 0.5, and tend to be less sensitive to choice of threshold method (Freeman and Moisen, 2008b; Liu et al., 2005; Liu et al., 2016). The analyses here illustrated that the optimal thresholds of species tended to be related to niche properties and model performance. When the 2 degree method was used to select pseudo-absences for CT and RT, optimal thresholds were high for species with restricted distribution ranges and/or high model accuracy (given the relationship between accuracy measures, Fig. B3; Table B2 in Mendeley Data). Species traits and model accuracy were also correlated with the sensitivity of species to threshold methods. For the 2 degree method, the difference in optimal thresholds among threshold methods was larger for specialist species than for generalist ones (Table B3 in Mendeley Data), and the spatial similarity in binary maps was high for species with high model accuracy and restricted ranges (Table B4 in Mendeley Data). Therefore, biological species traits can substantially influence the determination of optimal thresholds. In the present study, the default 0.5 was the worst one of the eight threshold methods. Its failure may be attributed to the lack of ecological reasoning and frequentist assumptions not aware what recursive partitioning trees and bagging bring to the table. Given that different species might require different thresholds (Table B2 in Mendeley Data), it was not surprisingly that the default 0.5 has been shown to perform poorly.

To make sure that the results were not influenced by the ubiquitous error sources from actual species data, researchers have often

developed geographically distinct virtual species (i.e. underlying mechanisms that generate these patterns are known) to emulate the characteristics of real world species (e.g. Barbet-Massin et al., 2012; Liu et al., 2013). Developing virtual species usually consists of two key steps: define species' environmental suitability from a spatial set of environmental variables, then use a probabilistic or threshold approach to convert environmental suitability to presence–absence map. For later, however, choice of conversion strategy often has significant consequences for SDM performance and optimal threshold methods selection (Meynard & Kaplan, 2013). This might be a reason why the performance of the threshold methods in our analysis were different from that in other studies (Liu et al., 2013). In addition, the most obvious fundamental drawback of these virtual species is the intrinsic lack of broader ecological processes, which are collectively a necessary component of determining species real-world distribution. Therefore, the methods widely accepted to develop virtual species are not available.

In line with the recent gradient theory (Cushman and Huettmann, 2010; Ter Braak and Prentice, 1988), however, providing numerical rather binary prediction maps is the most common method used in multiple and somewhat conflicting management applications (Freeman and Moisen, 2008b). That is because numerical results convey more information than binary outputs (Guillera-Arroita et al., 2015; Liu et al., 2013). In this manner, map users can choose appropriate threshold cut-off values and generate binary maps according to the intended map use (e.g. species range estimation, Kandel et al., 2015). For map making, unless sound justification exists for choosing a particular threshold cut-off over the others (e.g. a good data match, a high sensitivity is needed in defining a management area for a rare species, and a high specificity is needed for determining whether a species is threatened; Fielding and Bell, 1997; Liu et al., 2005, Nenzén and Araújo, 2011; Norris, 2014), there might be advantages in applying these four aforementioned threshold methods for CT and RT. Spatial uncertainty analysis revealed that a substantial difference exists in binary maps between CT and RT, and RT is more sensitive to choice of the threshold method than CT. Given the difference in spatial prediction among threshold approaches, we recommend to stay with Breiman (2001b) and infer from predictions. Further, we think ensemble models are a good solution and to apply these four threshold methods within an ensemble forecasting framework (Araújo and New, 2007) and to explore the resulting range of spatial uncertainties.

The present analysis demonstrated that the relative performance of CT and RT was different between the SRE and 2 degree methods in terms of model accuracy and spatial uncertainty in binary predictions (Figs. 2 and 4). Nevertheless, note that when using the RF model, we obtained a classifier or an estimate of the regression function, which was a piecewise constant function obtained by partitioning the predictor's space (Strobl et al., 2009). The SRE method also has a piecewise constant function in nature. Hence, it was not surprising that the SRE method performed better when compared with other pseudo-absence selection approaches in improving the SDMs (e.g. RF, GBM and CART) which inherently use tree based step-functions (Barbet-Massin et al., 2012). The fact that model structure rationality is more important than model accuracy gives us confidence that the adequacy of the SRE model maybe misleading for pseudo-absence selection for RF.

5. Conclusions

In conclusion, randomForest can perform as a leading prediction algorithm when used with multi-species and on a national level. However, in practice we argue for choosing RT rather than CT as the SDM if model discrimination capacity is viewed as more important than model reliability, and vice versa. In line with gradient theory, we recommend the use of probabilistic predictions of RT or CT for species distribution modeling. A binary conversion of model outputs should only be implemented when it is clearly justified by the application's objective. MaxKappa, MaxOA, MaxTSS, and MinROCdist are four promising objective methods recommended for binary conversion for CT and RT methods, while the CT default classification method (default 0.5) was not recommended for binary conversions. RT is more sensitive to the choice of threshold approach than CT. In addition, species with restricted ranges are most sensitive to the choice of threshold approaches in terms of optimal threshold values, whereas species with low model accuracy or wide ranges are most sensitive to the choice of threshold approaches in terms of spatial uncertainty in binary maps. We also recommend using the 2 degree method rather than the SRE method for selection of pseudo-absences for RF. These findings, coupled with the study showing how, where, and how many pseudo-absences should be generated for RF (Barbet-Massin et al., 2012), could help those who propose guidelines for the application of RF models in the field of species distribution modeling. We think our findings are generalizable and matter for wider applications world-wide.

Acknowledgements

This study was funded by the National Key R&D Program of China (2017YFC0505501, 2017YFC0505603) and National Natural Science Foundation of China (41301056).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2019.05.003.

References

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47.

Ballings, M., Van-den-Poel, D., 2013. AUC: Threshold Independent Performance Measures for Probabilistic Classifiers. R Package Version 0.3.0. https://CRAN.R-project.org/package=AUC.

Baltensperger, A.P., Morton, J.M., Huettmann, F., 2017. Expansion of American marten (*Martes americana*) distribution in response to climate and landscape change on the Kenai peninsula, Alaska. J. Mammal. 98, 703–714.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3, 327–338.

Bleyhl, B., Sipko, T., Trepet, S., Bragina, E., Leitão, P.J., Radeloff, V.C., Kuemmerle, T., 2015. Mapping seasonal European bison habitat in the Caucasus Mountains to identify potential reintroduction sites. Biol. Conserv. 191, 83–92.

Booth, T.H., Nix, H.A., Busby, J.R., Hutchinson, M.F., 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Divers. Distrib. 20, 1–9.

Breiman, L., 2001a. Random forests. Mach. Learn. 45, 5–32.

Breiman, L., 2001b. Statistical modeling: the two cultures. Stat. Sci. 16, 199–231.

Breiman, L., Cutler, A., 2004. Random Forests. http://www.math.usu.edu/~adele/forests/ (Cited 3 Nov, 2018).

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman & Hall/CRC, New York.

Brieuc, M.S.O., Waters, C.D., Drinana, D.P., Naish, K.A., 2018. A practical introduction to random forest for genetic association studies in ecology and evolution. Mol. Ecol. Resour. 18, 755–766.

Bryan, B.A., Gao, L., Ye, Y., Sun, X., Connor, J.D., Crossman, N.D., Stafford-Smith, M., Wu, J., He, C., Yu, D., Liu, Z., Li, A., Huang, Q., Ren, H., Deng, X., Zheng, H., Niu, J., Han, G., Hou, X., 2018. China's response to a national land-system sustainability emergency. Nature 559, 193–204.

Cao, Y., DeWalt, R.E., Robinson, J.L., Tweddale, T., Hinz, L., Pessino, M., 2013. Using Maxent to model the historic distributions of stonefly species in Illinois streams: the effects of regularization and threshold selections. Ecol. Model. 259, 30–39.

Cooper, J.C., Soberón, J., 2018. Creating individual accessible area hypotheses improves stacked species distribution model performance. Glob. Ecol. Biogeogr. 27, 156–165.

Craig, E., Huettmann, F., 2009. Using "blackbox" algorithms such as TreeNet and random forests for data-mining and for finding meaningful patterns, relationships and outliers in complex ecological data: An overview, an example using golden eagle satellite data and an outlook for a promising future. In: Wang, H.-F. (Ed.), Intelligent Data Analysis: Developing New Methodologies through Pattern Discovery and Recovery. IGI Global, Hershey, PA, USA, pp. 65–83.

Crimmins, S.M., Dobrowski, S.Z., Greenberg, J.A., Abatzoglou, J.T., Mynsberge, A.R., 2011. Changes in climatic water balance drive downhill shifts in plant species'

optimum elevations. Science 331, 324–327.
Cushman, S.A., Huettmann, F., 2010. Spatial Complexity, Informatics, and Wildlife Conservation. Springer, Tokyo, Japan.
Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.
De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.
Drew, C.A., Wiersma, Y.F., Huettmann, F., 2011. Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications. Springer, New York.
Editorial Board of Vegetation map of China (EBVMC), Chinese Academy of Sciences, 2001. 1:1,000,000 Vegetation Distribution Map of China. Science Press, Beijing.
Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.
Evans, J.S., Murphy, M.A., Holden, Z.A., Cushman, S.A., 2011. Modeling species distribution and change using random forest. In: Drew, C.A., Wiersma, Y.F., Huettmann, F. (Eds.), Predictive Species and Habitat Modeling in Landscape Ecology. Springer, New York, pp. 139–159.
FAO, 2015. Global Forest Resources Assessment 2015: Desk Reference. Food and Agriculture Organization of the United Nation, Rome, pp. 253.
Feller, W., 1968. An Introduction to Probability Theory and its Application, 3rd ed. vol. 1 Willey, New York.
Fernandes, R.F., Daniel, S., Antoine, G., 2018. How much should one sample to accurately predict the distribution of species assemblages? A virtual community approach. Ecol. Inform. 48, 125–134.
Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems. J. Mach. Learn. Res. 15, 3133–3181.
Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.
Freeman, E.A., Moisen, G., 2008a. PresenceAbsence: an R package for presence-absence model analysis. J. Stat. Softw. 23, 1–31.
Freeman, E.A., Moisen, G.G., 2008b. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and Kappa. Ecol. Model. 217, 48–58.
Gavish, Y., Marsh, C.J., Kuemmerlen, M., Stoll, S., Haase, P., Kunin, W.E., 2017. Accounting for biotic interactions through alpha-diversity constraints in stacked species distribution models. Methods Ecol. Evol. 8, 1092–1102.
Graham, C., Ferrier, S., Huettman, F., Moritz, C., Peterson, A., 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol. Evol. 19, 497–503.
Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R., Wintle, B.A., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Glob. Ecol. Biogeogr. 24, 276–292.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186.
Han, X., Huettmann, F., Guo, Y., Mi, C., Wen, L., 2018. Conservation prioritization with machine learning predictions for the black-necked crane Grus nigricollis, a flagship species on the Tibetan plateau for 2070. Reg. Environ. Chang. https://doi.org/10.1007/s10113-018-1336-4.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. Springer, New York.
Herrick, K.A., 2013. Predictive Modeling of Avian Influenza in Wild Birds. PhD thesis. University of Alaska-Fairbanks, Fairbanks, Alaska.
Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93, 679–688.
Huettmann, F., Ickert-Bond, S.M., 2018. On open access, data mining and plant conservation in the circumpolar north with an online data example of the herbarium, University of Alaska Museum of the North Arctic science. Sci. Arctique 4, 433–470. https://doi.org/10.1139/as-2016-0046. http://www.nrcresearchpress.com/toc/as/0/ja.
Ishwaran, H., Kogalur, U.B., 2007. Random Survival Forests for R. R News. vol. 7. pp. 25–31.
Jafariana, Z., Kargarb, M., Bahreini, Z., 2019. Which spatial distribution model best predicts the occurrence of dominant species in semi-arid rangeland of northern Iran? Ecol. Inform. 50, 33–42.
Jiménez-Valverde, A., Lobo, J., 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecol. 31, 361–369.
Kandel, K., Huettmann, F., Suwal, M.K., Regmi, G.R., Nijman, V., Nekaris, K.A.I., Lama, S.T., Thapa, A., Sharma, H.P., Subedi, T.R., 2015. Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: red panda (*Ailurus fulgens*) in the Hindu-Kush Himalaya region. Biol. Conserv. 181, 150–161.
Li, W., 2004. Degradation and restoration of forest ecosystems in China. For. Ecol. Manag. 201, 33–41.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.
Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28, 385–393.
Liu, C., White, M., Newell, G., 2011. Measuring and comparing the accuracy of species distribution models with presence–absence data. Ecography 34, 232–243.
Liu, C., White, M., Newell, G., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. J. Biogeogr. 40, 778–789.
Liu, C., Newell, G., White, M., 2016. On the selection of thresholds for predicting species occurrence with presence-only data. Ecol. Evol. 6, 337–348.
Mateo, R.G., Croat, T.B., Felicísimo, Á.M., Munoz, J., 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. Divers. Distrib. 16, 84–94.
Meynard, C.N., Kaplan, D.M., 2013. Using virtual species to study species distributions and model performance. J. Biogeogr. 40, 1–8.
Mi, C., Huettmann, F., Guo, Y., Han, X., Wen, L., 2017. Why choose random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. Peerj 5, e2849. https://doi.org/10.7717/peerj.2849.
Mueller, J.P., Massaron, L., 2016. Machine Learning for Dummies. John Wiley & Sons.
Nenzén, H., Araújo, M., 2011. Choice of threshold alters projections of species range shifts under climate change. Ecol. Model. 222, 3346–3354.
Norris, D., 2014. Model thresholds are more important than presence location type: understanding the distribution of lowland tapir (*Tapirus terrestris*) in a continuous Atlantic forest of Southeast Brazil. Trop. Conserv. Sci. 7, 529–547.
Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I., Louzao, M., 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. Biol. Conserv. 156, 94–104.
Patsiou, T.S., Conti, E., Zimmermann, N.E., Theodoridis, S., Randin, C.F., 2014. Topo-climatic microrefugia explain the persistence of a rare endemic plant in the Alps during the last 21 millenni. Glob. Chang. Biol. 20, 2286–2300.
Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 133, 225–245.
Peters, J., Baets, B.D., Verhoest, N.E.C., Samson, R., Degroeve, S., Becker, P.D., Huybrechts, W., 2007. Random forests as a tool for ecohydrological distribution modelling. Ecol. Model. 207, 304–318.
Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., Guisan, A., 2012. Climatic niche shifts are rare among terrestrial plant invaders. Science 335, 1344–1348.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Regmi, G.R., Huettmann, F., Suwal, M.K., Nijman, V., Nekaris, K.A.I., Kandel, K., Sharma, N., Coudrat, C., 2018. First open access ensemble climate envelope predictions of Assamese macaque *Macaca assamensis* in south and south-East Asia: a new role model and assessment of endangered species. Endanger. Species Res. 36, 149–160.
Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929.
Sarre, S.D., MacDonald, A.J., Barclay, C., Saunders, G.R., Ramsey, D.S., 2013. Foxes are now widespread in Tasmania: DNA detection defines the distribution of this rare but invasive carnivore. J. Appl. Ecol. 50, 459–468.
Song, C., Zhang, Y., 2010. Forest cover in China from 1949 to 2006. In: Nagendra, H., Southworth, J. (Eds.), Reforesting Landscapes: Linking Pattern and Process. Springer, Dordrecht, the Netherlands, pp. 341–356.
State Forestry Administration of China (SFAC), 2010. The Forestry Action Plan to Address Climate Change. China Forestry Press, Beijing.
Strobl, C., Malley, J.D., Tutz, G., 2009. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. Psychol. Methods 14, 323–348.
Ter Braak, C.J.F., Prentice, I.C., 1988. A theory of gradient analysis. Adv. Ecol. Res. 18, 271–317.
Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD: a platform for ensemble forecasting of species distributions. Ecography 32, 369–373.
Zaniewski, A., Lehmann, A., Overton, J., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol. Model. 157, 261–280.
Zhang, L., Wang, L., Zhang, X., Liu, S., Sun, P., Wang, T., 2014. The basic principle of random forest and its applications in ecology: a case study of *Pinus yunnanensis*. Acta Ecol. Sin. 34, 650–659.
Zhang, L., Liu, S., Sun, P., Wang, T., Wang, G., Wang, L., Zhang, X., 2016. Using DEM to predict *Abies faxoniana* and *Quercus aquifolioides* distributions in the upstream catchment basin of the Min River in Southwest China. Ecol. Indic. 69, 91–99.